Jurnal TRANSFORMATIKA
Vol. 20, No.2 Januari 2023, pp. 1 - 10
P-ISSN: 1693-3656, E-ISSN: 2460-6731
https://journals.usm.ac.id/index.php/transformatika/

Optimization of the Stemming Technique on *Text preprocessing* President 3 Periods Topic

M. Ulil Albab[1], Yohana Karuniawati P[2], Mohammad Nur Fawaiq[3]
[1]STIKES Estu Utomo
Kab. Boyolali, 081234302890, e-mail: ulilalbab@stikeseub.ac.id
[2]SMK Negeri 1 Kismantoro
Kab. Wonogiri, 081335001601, e-mail: yohana.karunia@gmail.com
[3]Program Studi Magister Teknik Informatika Universitas Amikom Yogyakarta
Jl. Padjajaran Ring Road Utara, Condongcatur, Depok, Sleman, (0274) 884201, e-mail: nurfawaiq@students.amikom.ac.id

ARTICLE INFO

History of the article :
Received 29 Agustus 2022
Received in revised form 29 Desember 2022
Accepted 5 januari 2023
Available online 8 31 Januari 2023

Keywords:

Optimasi; teknik stemming'text preprocessing; 3 periode;

*** Correspondece:**

Telepon:
081234302890

E-mail:
ulilalbab@stikeseub.ac.id

ABSTRACT (10 PT)

Stemming merupakan suatu proses untuk menemukan kata dasar dari sebuah kata. Penelitian ini bertujuan untuk melakukan tahapan text pre-processing pada data twitter yang menyebutkan topik "Presiden 3 Periode", yaitu sebanyak 797 data yang didapatkan dari crawling twitter mulai tanggal 15 April 2022 sampai dengan 30 April 2022, sekaligus melakukan optimasi salah satu teknik stemming terhadap teks berbahasa Indonesia yang memang belum banyak dilakukan. Banyaknya kata yang diolah sebanyak 9401 kata. Optimasi yang dilakukan yaitu dengan memodifikasi kamus bahasa dan dengan menambahkan kata-kata yang dimasukkan dalam *Stopword* sehingga dapat menghasilkan jumlah kata ter-stemming yang semakin meningkat. Sebelum dilakukan optimasi, prosentase keberhasilan stemming mencapai 95,86%, setelah dilakukan optimasi meningkat menjadi 99,93%

1. INTRODUCTION (Bold, 11 PT)

Twitter merupakan media yang cukup banyak digunakan oleh masyarakat Indonesia dalam menyampaikan aspirasinya, yaitu sebanyak 18,45 juta pengguna pada awal 2022[13] termasuk dalam hal berpendapat terkait politik di Indonesia. Wacana ini terus bergulir setelah disuarakan oleh sejumlah elit partai maupun menteri di kabinet. Akan tetapi, menjelang berakhirnya masa jabatan Presiden Jokowi pada periode kedua ini, muncul isu-isu mengenai Presiden yang dipilih lagi sehingga muncul topik Presiden 3 Periode. Masyarakat banyak menggunakan topik ini di berbagai media sosial, seperti Twitter, Instagram dan lain sebagainya. Dari berbagai opini masyarakat yang muncul, banyak yang bernada sentimen. Sentimen itu sendiri adalah sebuah perasaan yang menggambarkan sesuatu yang bersifat bisa positif, negatif dan netral. Biasanya para

peneliti melakukan observasi pada opini - opini masyarakat guna diketahui apakah sentimen masyarakat tersebut termasuk kedalam kategori positif, negatif dan netral.

Sebelum dilakukan pengkategorian sentimen ada tahapan preprocessing yang harus dilalui terlebih dahulu, salah satunya adalah stemming. Stemming adalah proses mengembalikan kata kedalam bentuk dasar. Jika stemming berhasil dilakukan maka akan menghasilkan klasifikasi sentimen yang akurat. Akan tetapi adakalanya stemming itu masih belum berhasil karena ada beberapa kata yang seharusnya tidak di-stemming namun tetap di-stemming [8]. Sebagai contoh penelitian yang dilakukan oleh A. Yudi Permana dkk, bahwa kesalahan karena algoritma Porter-KBBI terjadi ketika sebuah kata tidak ditemukan dalam kamus basis data kemudian dianggap sebagai kata dasar, dan kesalahan stemming terhadap nama orang yang seharusnya tidak dilakukan stemming [10]. Maka dari itu perlu dilakukan optimasi terhadap proses stemming sehingga menghasilkan luaran stemming yang lebih optimal. Tim peneliti mulai melakukan pengumpulan data dan mulai melakukan *Text preprocessing* pada data yang telah dikumpulkan. Stemming kata untuk teks bahasa Indonesia berbeda dengan stemming kata untuk teks bahasa Inggris. Untuk teks bahasa Inggris, satu-satunya proses yang diperlukan adalah menghilangkan akhiran. Di sisi lain, teks bahasa Indonesia juga menghilangkan semua sufiks, baik sufiks maupun prefiks..

Adapun tujuan dari penelitian ini adalah untuk mengoptimalkan tahapan stemming dalam proses *Text preprocessing* pada data yang diunduh dari Twitter dengan topik Presiden 3 Periode. Sedangkan untuk rumusan masalah yaitu bagaimana mengoptimalkan tahapan stemming pada tahapan *Text preprocessing* pada data yang diunduh dari Twitter dengan topik Presiden 3 Periode. Untuk pembatasan masalah, tim peneliti menetapkan bahasan yaitu kata yang dilakukan dalam tahapan *Text preprocessing* adalah stemming berbahasa Indonesia, kata-kata baku Bahasa Indonesia, tidak meliputi kata yang disingkat, dan tidak meliputi kata-kata bahasa gaul yang tidak sesuai EYD serta media sosial yang digunakan untuk pengambilan data adalah media sosial twitter.

Untuk mendasari dilakukannya penelitian ini, maka diperlukan penelitian sebelumnya atau penelitian terkait. Karena penelitian yang bagus adalah yang berlanjut dari sebelumnya atau yang mempunyai dasar dari penelitian sebelumnya. Yudi Permana melakukan penelitian untuk mencari nilai akurasi klasifikasi topik soal UN. Metode penelitian yang digunakan yaitu stemming porter dan diperoleh akurasi data training 94,34% dan data testing 72.67%[11]. Bunga Chintia melakukan penelitian untuk mengukur efektifitas penggunaan beberapa parameter pada algoritma stemming, seperti akurasi dan waktu proses serta dilakukan tahap uji stemming dengan suatu corpus. Hasil yang dicapai yaitu 87% adalah algoritma Arifin dan Setiono[12]. Rahardyan melakukan penelitian dengan Levenshtein Distance dan berhasil mendapatkan tingkat akurasi 96.6%, hasil pengujian algoritma Non-formal Affix mendapatkan akurasi 73.3%[14].

RESEARCH METHODS (Bold, 11 PT)

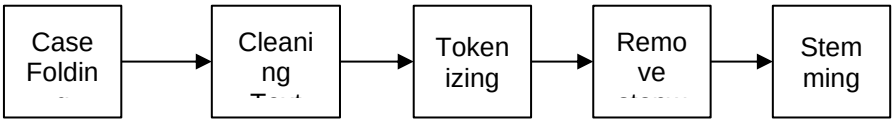
1. Scrapping Data

Scraping Data adalah sebuah cara komputer untuk melakukan penguraian intisari informasi dari sebuah data [2]. Pada tahap ini dilakukan pengambilan data dari Twitter menggunakan kata kunci “Presiden 3 Periode”, yang dimulai dari 15 April 2022 sampai 30 April 2022 sebanyak 797 data. Data yang telah diambil kemudian disimpan kedalam format csv. Data ini adalah data yang bersifat mentah dan belum siap untuk dilakukan penelitian. Data mentah tersebut diberikan nama atau label sesuai kebutuhan penelitian. Variabel yang dibentuk dan dibutuhkan pada penelitian ini adalah variabel tweet dan tanggal tweet itu dibuat.

2. Text Preprocessing

Menurut Oueslati et al [3], *Text preprocessing* merupakan tahapan yang sangat penting sebelum langkah awal dimulainya suatu penelitian. Karena suatu penelitian dikatakan berhasil dan lancar jika dalam text preproccessingnya sangat minim kesalahan. *Text preprocessing*

dilakukan dengan tujuan dimana data awal diproses dengan melewati beberapa tahapan hingga data tersebut benar-benar siap untuk digunakan. Gambar 2.1 menunjukkan tahapan *Text preprocessing* yang dilakukan dalam penelitian ini.



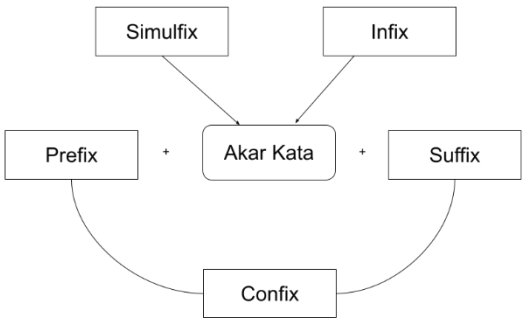
Gambar 1. Tahapan Text Preprocessing

Case folding merupakan suatu tahapan proses untuk mengubah kata menjadi bentuk yang sama menggunakan Python string lower method [4]. Tujuan dari *Case folding* adalah mengembalikan semua kata kedalam bentuk huruf kecil semua supaya data text yang diproses semuanya dalam kondisi bentuk yang sama. Cleaning teks merupakan proses pembersihan validasi kata yang tidak diinginkan untuk mengurangi gangguan pada proses klasifikasi. Kata yang dihilangkan misalnya karakter.

Tokenisasi adalah metode membagi secara kasar serangkaian karakter dalam teks menjadi kata-kata untuk membedakan antara karakter tertentu yang mungkin atau mungkin tidak diperlakukan sebagai jeda kata. Contoh kata break adalah karakter spasi putih seperti Enter, Tabulator, dan spasi. Namun, tanda kutip tunggal (`), titik (.), Titik koma (;), titik dua (:), dll., memiliki banyak peran sebagai pemisah kata. Bagaimana karakter dalam teks diperlakukan tergantung pada konteks aplikasi yang sedang dikembangkan. Tugas tokenization ini menjadi lebih sulit jika Anda juga harus memperhatikan struktur (tata bahasa) bahasa tersebut. [8]. Kemudian dilakukan proses remove *Stopword*, yaitu proses filtering, pemilihan kata-kata penting dari hasil token yaitu kata-kata apa saja yang digunakan untuk mewakili dokumen[16]. Selanjutnya dilakukan stemming. Stemming adalah proses pemetaan dan penguraian berbagai bentuk kata menjadi bentuk dasarnya[15]. Proses pemetaan dan penguraian digunakan untuk menemukan kata dasar dari sebuah kata yang mengalami imbuhan dengan cara menghilangkan atau menghapus imbuhan-imbuhan tersebut.

3. Struktur Imbuhan Bahasa Indonesia

Berdasarkan penelitian sebelumnya, struktur sufiks bahasa Indonesia terdiri dari prefiks, sufiks, infiks, konfiks, dan simulfiks. [1]. Struktur imbuhan pada Bahasa Indonesia dapat dilihat pada Gambar 2



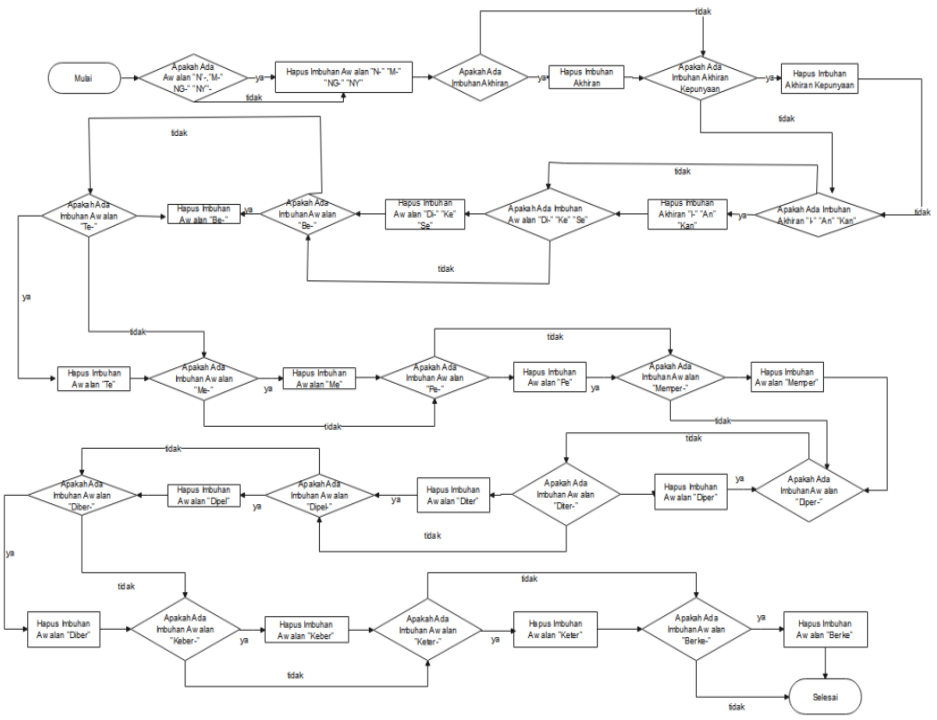
Gambar 2 Struktur Imbuhan Bahasa Indonesia

DOI : http://dx.doi.org/10.26623/transformatika.v20i2.5374

Prefix/awalan adalah imbuhan yang berada di awal sebuah akar kata, antara lain me-, di- , ter-, pe-, se-. Contoh kata dengan prefix : melihat, dibaca, terdiam. Suffix/akhiran adalah imbuhan yang diletakkan di belakang akar kata, antara lain -an,-i,-kan. Contoh : sayuran, akhiri, letakkan. Infix/sisipan adalah imbuhan yang dibubuhkan pada tengah-tengah kata antara lain -el-, -er-, -em-. Contoh : selidik, serabut, cemerlang. Simulfiks adalah afiks yang menggantikan huruf di depan suatu kata. Contoh : mencuci menjadi nyuci, mencari menjadi nyari. Konfiks merupakan kombinasi dari prefiks dan sufiks untuk membentuk kata baru yang berasal dari kata dasar, antara lain meng- -kan, di- -kan, per- -an. Contoh : menghabiskan, dijauhkan, persaingan [14].

4. Algoritma Nazief & Andriani

Pada penelitian sebelumnya[11] dikembangkan algoritma yang didasarkan pada algoritma Nazief & Andriani. Pengembangan yang dilakukan untuk meningkatkan akurasi algoritma Nazief & Andriani untuk melakukan ekstraksi kata dasar pada kata berimbuhan tidak baku. Gambar 3 menunjukkan flowchart stemming pada algoritma Nazief & Andriani.

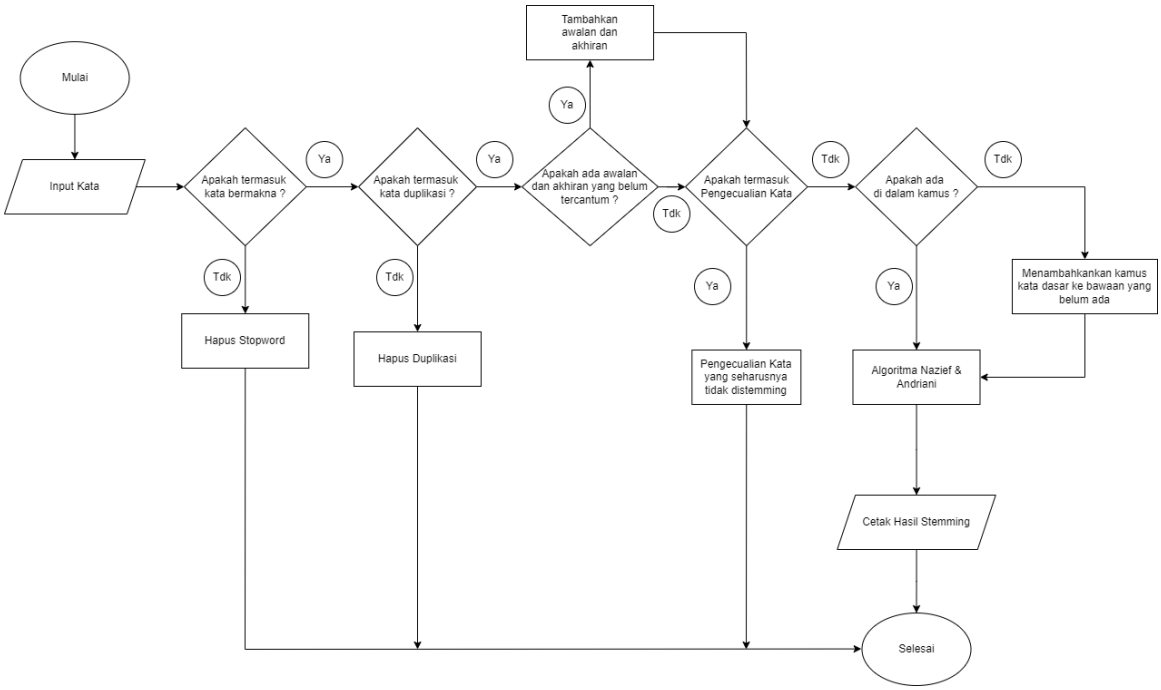


Gambar 3 Flowchart Stemming Nazief & Andriani

5. Optimasi Teknik Stemming

Didalam melakukan stemming seringkali dijumpai hasil stemming yang kurang maksimal. Oleh karena itu penelitian ini akan memaksimalkan stemming sehingga menghasilkan olahan stemming yang paling baik. [9] Sebagai langkah awal pada stemming, digunakan suatu library khusus untuk stemming pada bahasa pemrograman python supaya memudahkan dalam mengimplementasikannya. Library khusus yang digunakan adalah library

sastrawi dan NLTK dengan algoritma Nazief dan Andriani didalamnya. Cara kerja algoritma ini adalah setiap kata yang ada dilakukan proses stemming. Jadi sebenarnya tidak semua kata diharuskan untuk di stemming. Oleh karena itu ada modifikasi algoritma pada proses stemming ini, sehingga luaran hasil yang didapatkan akan lebih optimal. Adapun algoritma modifikasi untuk optimasi stemming ditunjukkan pada Gambar 4



Gambar 4, Flowchart Optimasi Teknik Stemming

a. Library NLTK

NLTK merupakan rangkaian perpustakaan dan program pengolahan bahasa simbolik dan statistic alami untuk pemrosesan teks bahasa inggris yang ditulis dalam bahasa Pemrograman Python [9]. Mudahnya, NLTK merupakan library khusus pendukung untuk pengolahan text stemming yang menggunakan bahasa pemrograman python. Dengan adanya library ini seorang peneliti akan dengan mudah melakukan stemming karena hanya perlu memanggil fungsi dan sintax librarynya saja.

b. Library Sastrawi

Salah satu kekurangan NLTK adalah dukungan bahasa Indonesia yang masih kurang. Oleh karena itu, kami akan menggunakan perpustakaan tambahan berupa literatur. Sastrawi adalah perpustakaan NLP yang didedikasikan untuk bahasa Indonesia. Awalnya, Library dikembangkan dan ditujukan untuk bahasa pemrograman PHP, tetapi karena popularitasnya, perpustakaan ini juga dikembangkan untuk mendukung bahasa pemrograman Python. [7].

c. Kamus

Pada tahap ini dilakukan pemberian skor dengan didasarkan pada kamus lexicon. Kamus leksikon berisi kata-kata opini untuk menilai kelas emosi positif dan negatif dari data tekstual. Kamus penelitian ini menggunakan kamus opini positif dan negatif yang diterjemahkan ke dalam bahasa Indonesia dari penelitian Liu et al., (2005). Perhitungan

DOI : http://dx.doi.org/10.26623/transformatika.v20i2.5374

nilai label emosi menggunakan metode penghitungan jumlah skor emosi kelas positif dikurangi dengan skor emosi kelas negatif setiap komentar. [17].

d. *Stopword*

Stopword ini adalah kata-kata yang dihapus dan tidak akan digunakan dalam proses stemming. Dalam penelitian ini, beberapa *Stopword* seperti sih, lah, lu, aja.

RESULTS AND DISCUSSION

In this section, it is explained the results of research and at the same time is given the comprehensive discussion. Results can be presented in figures, graphs, tables and others that make the reader understand easily.

Proses modifikasi pada algoritma stemming untuk optimasi hasil stemming dilakukan guna mengatasi kesalahan dari stemming. Serangkaian proses tahapannya adalah dimulai dari cleaning text menghapus kata-kata yang tidak diinginkan terlebih dahulu atau tidak digunakan, sehingga akan dihapus pada proses ini. Tabel 1 menunjukkan beberapa kata yang dihapus.

Tabel 1. *Stopword*

CONTOH *STOPWORD*							
'ada'	'adalah'	'adanya'	'adapun'	'agak'	'agaknya'	'agar'	'akan'
'akankah'	'akhir'	'akhiri'	'akhirnya'	'aku'	'akulah'	'amat'	'amatlah'
'anda'	'andalah'	'antar'	'antara'	‘antaranya'	'apa'	'apaan'	'apabila'
'belum'	'belumlah'	'benar'	'benarkah'	'benarlah'	'berada'	'berakhir'	‘berakhirlah'
'cara'	'caranya'	'cukup'	cukupkah'	'cukuplah'	'cuma'	'dahulu'	'dalam'

Kemudian proses selanjutnya adalah menghapus duplikasi kata pada sebuah tweet. Sebelum menghapus duplikasi, ada lebih dari 10.000 kata, setelah dilakukan hapus duplikasi, tersisa 9401 kata untuk dilakukan proses stemming. Kemudian optimasi juga dilakukan dengan menambahkan awalan dan akhiran di proses stemming yg belum ada, seperti pada Tabel 3

Tabel 3. Penambahan awalan dan akhiran

	(Nazief-Andriani)	**Optimasi**
Prefix	di-, ke-, se-	di-, ke-, se-, peng-, nge-, me-, men-
Suffix	-i, -kan, -an, -is, -isme, -isasi	-i, -kan, -an, -is, -isme, -isasi, -in, -isir

Dalam melakukan sebuah analisis untuk mengukur ketepatan atau keberhasilan stemming, maka perlu dilakukan analisis secara manual tiap kata perkata. Dari proses ini akan terlihat mana yang terjadi pada kesalahan proses stemming dan mana yang berhasil dilakukan stemming. Kesalahan stemming terjadi karena kata yang hendak diproses stemming tidak ada di dalam kamus, sehingga menghasilkan luaran stemming yang tidak tepat. Beberapa kata yang ada di dalam kamus

ditunjukkan pada Tabel 4. Kesalahan hasil stemming untuk kata yang tidak tepat pada algoritma Nazief & Andriani ditunjukkan pada Tabel 5

Tabel 4. Kamus Kata Dasar Algoritma *Nazief-Andriani*

aba-aba	koalisi	substansial
abad	koana	substantif
advis	koar	substitusi
advokasi	kobah	substitutif
aerasi	kobak	substrat
tombak	kobalamin	subtil
tombok	kobalt	subtonik
tombol	kobar	subtropik
zuriah	kober	subuco
zus	koboi	subuh
pendapat	koboisme	subunit
sesuai	kobok	subur

Tabel 5. Kesalahan Stemming Algoritma Nazief & Andriani

Kata	Seharusnya	Terdeteksi	Status
mentri	mentri	tri	Salah stemming
pemilu	pemilu	milu	Salah stemming
ketumnya	ketum	ketumnya	Salah stemming
gerakan	gerak	gera	Salah stemming
capresnya	capres	capresnya	Salah stemming
gini	gini	gin	Salah stemming
milan	milan	mil	Salah stemming
mencengkram	cengkram	mencengkram	Salah stemming
penghadang	hadang	penghadang	Salah stemming
dicapreskan	capres	dicapreskan	Salah stemming
jagain	jaga	jagain	Salah stemming

DOI : http://dx.doi.org/10.26623/transformatika.v20i2.5374

Kata	Seharusnya	Terdeteksi	Status
menikah	nikah	meni	Salah stemming
rosi	rosi	ros	Salah stemming
pengertiannya	mengerti	erti	Salah stemming
setuju	setuju	tuju	Salah stemming
menetralisir	netral	menetralisir	Salah stemming
kulik	kulik	lik	Salah stemming

Kemudian kesalahan stemming juga terjadi untuk kata-kata yang seharusnya tidak diproses untuk stemming namun tetap terproses stemming. Tabel 6 menunjukkan kesalahan proses stemming karena kata yang seharusnya tidak di proses untuk stemming namun ikut terstemming.

Tabel 6. Kesalahan Stemming karena kata yang tidak harus di stemming

Kata	Seharusnya
mentri	Tidak perlu stemming
pemilu	Tidak perlu stemming
ketum	Tidak perlu stemming
gini	Tidak perlu stemming
milan	Tidak perlu stemming
rosi	Tidak perlu stemming
setuju	Tidak perlu stemming

Untuk mengatasi kesalahan tersebut, dilakukan optimasi stemming dengan penambahan pada kamus kata dasar. Optimasi yang dilakukan yaitu dengan menambahkan kata pada Tabel 7 pada kamus kata dasar algoritma Nazief-Andriani. Kamus kata dasar pada algoritma Nazief-Andriani memiliki 29.932 kata. Pada penelitian ini, peneliti menambahkan menjadi 29.938 kata.

Tabel 7. Penambahan Kata pada Kamus Kata Dasar Algoritma Nazief-Andriani

Daftar Kata	
ketum	setuju
capres	cengkram
hadang	mengerti

Optimasi juga dilakukan dengan membuat daftar kata pengecualian, agar kata yang berada di dalam daftar tersebut tidak ikut terstemming. Tabel 8 menunjukkan daftar kata pengecualian dan banyaknya tweet dengan kata tersebut yang berhasil dioptimasi. Tabel 9 menunjukkan data mentah yang sudah melalui tahapan text preprocessing sampai dengan optimasi stemming.

Tabel 8. Daftar Kata Pengecualian

Kata Pengecualian	Banyak tweet
mentri	3
kendaraan	1
pemilu	204
gini	4
milan	1
perhatian	16
rosi	1
setuju	1
kulik	1

Tabel 9. Hasil optimasi stemming dari data mentah

Tweet	Case folding	Cleaning	Non Stop word	Stemming	Optimasi stemming
b'Presiden Jokowi memang sih sudah perintahkan menterinya untuk setop wacana presiden 3 periode, tapi diam-diam ada bocoran informasi bahwa, dukungan presiden 3 periode bakal digas lagi saat momentum	b'presiden jokowi memang sih sudah perintahkan menterinya untuk setop wacana presiden 3 periode, tapi diam-diam ada bocoran informasi bahwa, dukungan presiden 3 periode bakal digas lagi saat momentum	presiden jokowi memang sih sudah perintahkan menterinya untuk setop wacana presiden periode tapi diamdiam ada bocoran informasi bahwa dukungan presiden periode bakal digas lagi saat momentum	['presiden', 'jokowi', 'perintahkan', 'menterinya', 'setop', 'wacana', 'presiden', 'periode', 'diamdiam', 'bocoran', 'informasi', 'dukungan', 'presiden', 'periode', 'digas', 'momentum', 'peringatan', 'lahir', 'pancasilan']	['presiden', 'jokowi', 'perintah', 'menteri', 'setop', 'wacana', 'presiden', 'periode', 'diamdiam', 'bocor', 'informasi', 'dukung', 'presiden', 'periode', 'gas', 'momentum', 'ingat', 'lahir', 'pancasilan']	['presiden', 'jokowi', 'perintah', 'menteri', 'setop', 'wacana', 'presiden', 'periode', 'diamdiam', 'bocor', 'informasi', 'dukung', 'presiden', 'periode', 'gas', 'momentum', 'ingat', 'lahir', 'pancasilan']

Tweet	Case folding	Cleaning	Non Stop word	Stemming	Optimasi stemming
Peringatan Hari Lahir Pancasila.\nhttps://t.co/ICrlRrl1oG'	peringatan hari lahir pancasila.\nhttps://t.co/icrlrrl1og'	peringatan hari lahir pancasilan			
b'Komitmen Presiden jelas ya, jika masih ada pihak yg bersuara penundaan pemilu/ jabatan 3 periode artinya mereka sendiri yg ingin bikin gaduh https://t.co/y631o6f93U'	b'komitmen presiden jelas ya, jika masih ada pihak yg bersuara penundaan pemilu/ jabatan 3 periode artinya mereka sendiri yg ingin bikin gaduh https://t.co/y631o6f93u'	komitmen presiden jelas ya jika masih ada pihak yg bersuara penundaan pemilu jabatan periode artinya mereka sendiri yg ingin bikin gaduh	['komitmen', 'presiden', 'bersuara', 'penundaan', 'pemilu', 'jabatan', 'periode', 'bikin', 'gaduh']	['komitmen', 'presiden', 'suara', 'tunda', 'milu', 'jabat', 'periode', 'bikin', 'gaduh']	['komitmen', 'presiden', 'suara', 'tunda', 'pemilu', 'jabat', 'periode', 'bikin', 'gaduh']
b'Copot mentri RI yang mendukung presiden 3 periode.!'	b'Copot mentri RI yang mendukung presiden 3 periode.!'	copot mentri ri yang mendukung presiden periode	['copot', 'mentri', 'ri', 'mendukung', 'presiden', 'periode']	['copot', 'tri', 'ri', 'dukung', 'presiden', 'periode']	['copot', 'mentri', 'ri', 'dukung', 'presiden', 'periode']

Untuk pengujian sebagai tolok ukur berhasilnya proses stemming dan optimasi stemming ini maka dilakukan evaluasi dengan menghitung prosentase berapa banyak kata yang berhasil di-stemming dibagi dengan jumlah total kata, kemudian dibandingkan manakah prosentase yang lebih tinggi apakah hasil stemming awal atau hasil stemming setelah optimasi yang memiliki prosentase lebih tinggi. Hasil yang didapatkan ditunjukkan pada tabel 10.

Tabel 10. Hasil prosentasi pengujian

	Algoritma Nazief-Andriani	Optimasi
Jumlah total kata	9401	9401
Berhasil di-stemming	9012	9395
Prosentase	95,86 %	99,93%

CONCLUSIONS AND RECOMMENDATIONS

Proses stemming menggunakan algoritma Nazief-Andriani cukup berhasil di beberapa penelitian. Tetapi pada penelitian kali ini, penggunaan algoritma Nazief-Andriani mendapatkan prosentase keberhasilan sebesar 95,86%. Kemudian dilakukan optimasi dengan menambahkan list stopword, menghapus duplikasi kata dalam satu cleaning tweet, menambahkan awalan dan akhiran di proses stemming, menambahkan kamus kata dasar, membuat list pengecualian untuk kata dasar yg seharusnya tidak ter-stemming tetapi ikut ter-stemming.

In this section, it can also be added the prospect of the development of research results and application prospects of further studies into the next (based on result and discussion).

REFERENCES

[1] R. B. S. Putra and E. Utami, "Non-formal affixed word stemming in Indonesian language," *2018 Int. Conf. Inf. Commun. Technol. ICOIACT 2018*, vol. 2018-January, pp. 531–536, 2018, doi: 10.1109/ICOIACT.2018.8350735.

[2] Nasution, Nana Nerina, "Sistem Pengumpulan Data Publikasi Ilmiah Menggunakan Web Crawling" Program Studi Teknologi Informasi, Universitas Sumatera Utara : 2020.

[3] Oumaima Oueslati, et al., A review of sentiment analysis research in Arabic language, Future Generat. Comput. Syst. (2020), 2020.

[4] Ronal Watrianthos, Samsir, Basyarul Ulya, Junaidi Mustapa Harahap, Deci Irmayani, Firman Edi, Jupriaman, Rizki Kurniawan Rangkuti (2021) , Naives Bayes Algorithm for Twitter Sentiment Analysis, https://iopscience.iop.org/article/10.1088/1742-6596/1933/1/012019/pdf

[5] Murnawan, M. (2017). Pemanfaatan Analisis Sentimen Untuk Pemeringkatan Popularitas Tujuan Wisata. *Jurnal Penelitian Pos dan Informatika*, *7*(2), 109-120. https://202.89.117.131/index.php/jppi/article/viewFile/070203/99

[6] E. J. Rifano, Abd. C. Fauzan, A. Makhi, E. Nadya, Z. Nasikin, and F. N. Putra, "Text Summarization Menggunakan Library Natural Language Toolkit (NLTK) Berbasis Pemrograman Python," ILKOMNIKA: Journal of Computer Science and Applied Informatics, vol. 2, no. 1, pp. 8–17, Apr. 2020, doi: 10.28926/ilkomnika.v2i1.32.

[7] A. Y. Permana and M. M. Effendi, "Optimasi Stemming Porter KBBI dan Cross Validation Naïve Bayes untuk Klasifikasi Topik Soal UN Bahasa Indonesia," *J. Ilm. Komputasi*, vol. 17, no. 4, 2018, doi: 10.32409/jikstik.17.4.2492.

[8] Rezalina, O. (2016). Perbandingan Algoritma Stemming Nazief & Adriani, Porter dan Arifin Setiono untuk Dokumen Teks Bahasa Indonesia (Doctoral Dissertation, Universitas Muhammadiyah Jember).<http://repository.unmuhjember.ac.id/550/1/JURNAL.pdf > (Diakses 13 Mei 2022)

[9] Ningrum, B. C. (2019). Perbandingan Algoritma Stemming untuk Bahasa Indonesia dengan Parameter Akurasi dan Waktu Proses. <https://repositori.usu.ac.id/bitstream/handle/123456789/23008/141401022.pdf?sequence=1&isAllowed=y > (Diakses 13 Mei 2022)

[10] Databoks Katadata. (2022). Pengguna Twitter Indonesia Masuk Daftar Terbanyak di Dunia, Urutan Berapa? https://databoks.katadata.co.id/datapublish/2022/03/23/pengguna-twitter-indonesia-masuk-daftar-terbanyak-di-dunia-urutan-berapa

[11] Putra, R. B. S., Utami, E., & Raharjo, S. (2018). Optimalisasi Stemming Kata Berimbuhan Tidak Baku Pada Bahasa Indonesia Dengan Levenshtein Distance. *Jurnal Informatika: Jurnal Pengembangan IT*, *3*(2), 200-205. http://ejournal.poltektegal.ac.id/index.php/informatika/article/download/877/696

[12] Amrullah, A. Z., Anas, A. S., & Hidayat, M. A. J. (2020). Analisis Sentimen Movie Review Menggunakan Naive Bayes Classifier Dengan Seleksi Fitur Chi Square. *Jurnal Bumigora Information Technology (BITe)*, *2*(1), 40-44. https://journal.universitasbumigora.ac.id/index.php/bite/article/download/804/527

[13] Anwar, M. S., Subroto, I. M. I., & Mulyono, S. (2020). Sistem Pencarian E-Journal Menggunakan Metode Stopword Removal dan Stemming Berbasis Android. *Prosiding Konstelasi Ilmiah Mahasiswa Unissula (KIMU) Klaster Engineering*. http://lppm-unissula.com/jurnal.unissula.ac.id/index.php/kimueng/article/download/8420/3887

[14] Mahendrajaya, R., Buntoro, G. A., & Setyawan, M. B, 2019, Analisis Sentimen Pengguna Gopay Menggunakan Metode Lexicon Based Dan Support Vector Machine. Komputek : Jurnal Teknik Universitas Muhammadiyah Ponorogo, No.2, Vol.3, 52–63. Doi: http://studentjournal.umpo.ac.id/index.php/komputek/article/view/270